



your dedicated on premise AI packaged solution





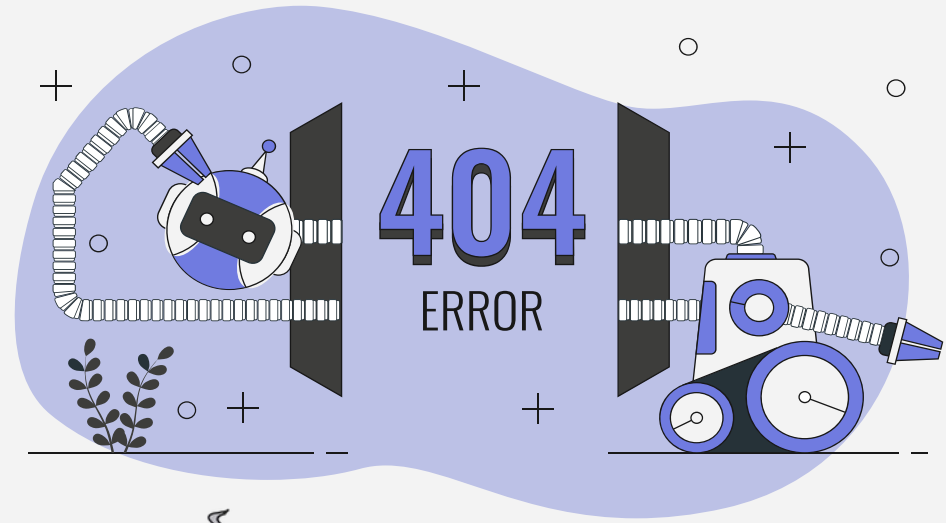
aiborg desktop is your complete on premise ready-to-use solution (hardware, software and expertise), for all your AI/voice/visio applications and interactions (prompt and more...).





# Pain points?

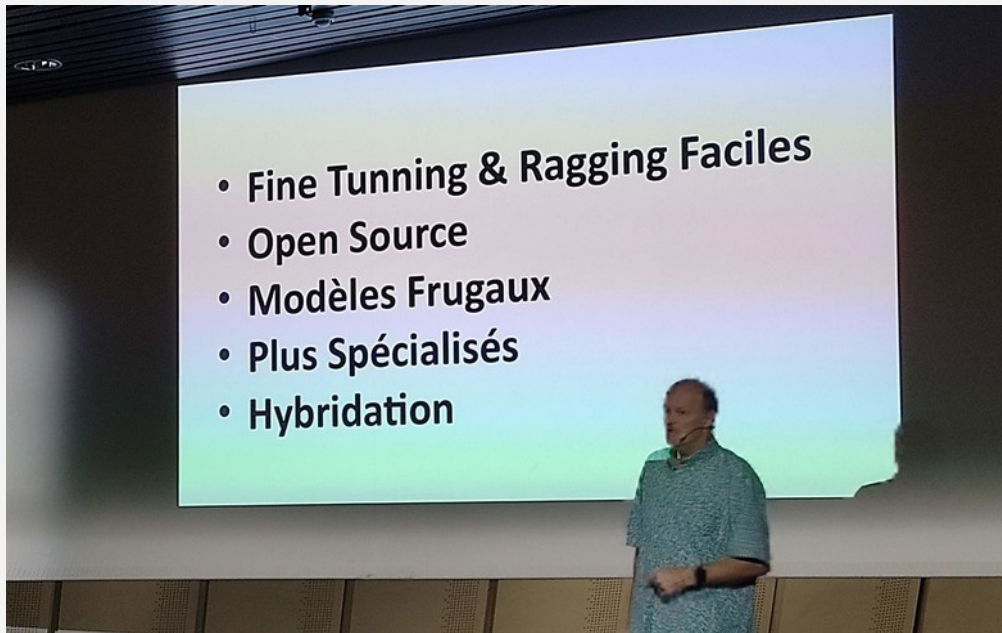
- Which Agent(s)
- Which LLM(s)
- RAG
- Tensors
- Contextes
- Token's
- Security
- Data
- Hardware
- Delays, poor achievements
- Latencies ...



# AIBORG



Our goals aligned with Luc key  
takeaway.



- Fine Tuning & Ragging Faciles
- Open Source
- Modèles Frugaux
- Plus Spécialisés
- Hybridation

Luc Julia - Chief Scientific Officer - Renault Group  
Conclusions of his keynote «Mythes et réalités des IA génératives  
et générales, et Fireside chat sur l'IA open source», #OSXP  
december 2024





# LLM + voice & visio immediately available

- Zero API/token costs, no usage limits;
- Video conferencing integration;
- Integration with realtime Chat/Visio/VoIP technologies;
- Docker technology without cloud;
- Private jitsi server with openfire.
- Ultra flexible for easy LLM changes;
- Immediately usable for simple functions;
- Translations, summaries, questions & answers;
- Basic voice assistant, internet search bot.



*Everything on site, functional and under control.*




# Company creation



Since 2006, Free-solutions sàrl is offering you a way to take a total control over all your interactions and discussions !

Provide close privileged collaboration with developers and experts.

Demonstrate a proven and established business model.

Provide Free Software, definitions from the  **FREE SOFTWARE**  
FOUNDATION

*To regain a total control of all your voice/visio/data interactions.*





## Company expertise



Free-Solutions Sàrl has been providing IT services based entirely on FREE solutions (with no external service providers or software).

On premise, SaaS or dedicated server for a total control of all your chat, audio, visio and other communications...

Free-solutions has been a driving force behind the **Chat XMPP Openfire** opensource project since its creation.



# ..... Company offers

## Free! Solutions



### **FREE and value-for-money solutions:**

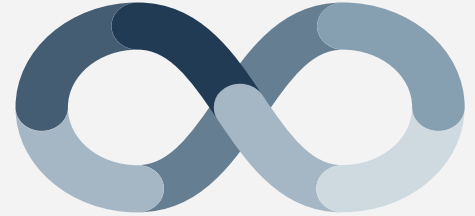
- No need third parties such as GAFAM's;
- No need CLOUD;
- FREE because your communications are secure and private;
- FREE because your data are not used neither managed by third parties;
- FREE because 100% of the code used is FREE;
- Sovereignty and eco-friendly.







# Why aiborg ?



Integrates AI naturally with chat/voice and therefore realtime technologies;

AI / XMPP / webRTC & simultaneous translations;

Combine webRTC with AI gives you the multimedia capabilities with LLM's;

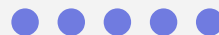
Lake of standard hardware offers on the market matching with our technical requirements;

Build our AI aiborg desktop in 2024, complete, powerful, ready to use and using all these technologies.

*For those who need LLM's and a complete AI environment on site.*



# aiborg desktop in brief



An AI work properly in inference mode (use),  
thank's to the right hardware:

- Powerful GPU: **32 GB** min
- Ultra-fast disks with 2x 4 **TB** capacity: **nvram**
- Memory for RAMDRIVES: **192GB** min
- CPU ryzen or I9: **32 cores**



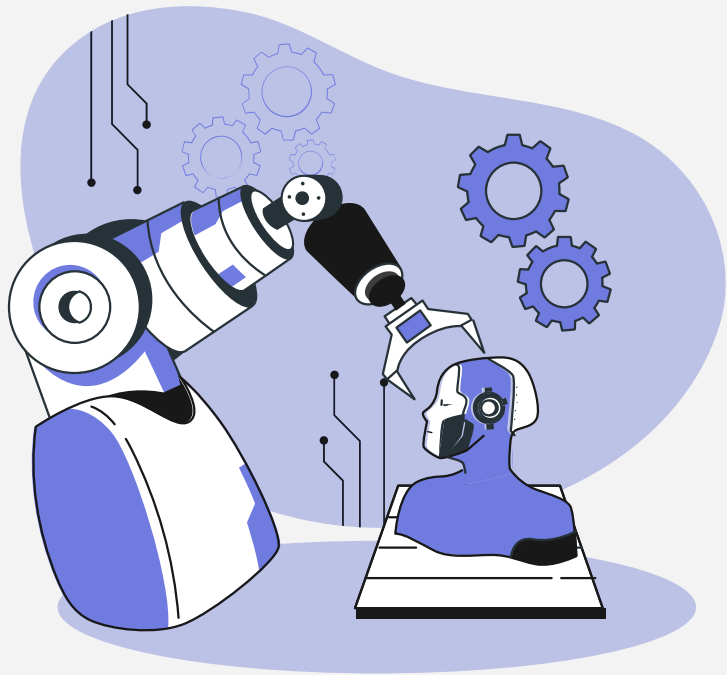
## ...on premise

### LAN/WAN/DMZ

- OS, kernel modules, kernel tuning, OS tuning,
- Compilation of C code, libcurl, ffmpeg, json, python,
- **Devops docker with GPU,**
- nvidia drivers installed and configured, source code
- and compilation of llama.ccp for Mistral...

Components	Supplier	Model	Qty
Processor 32 cores	Intel/AMD	i9/Ryzen	1
Motherboard	Gigabyte	B850+	1
Graphics card 32 GB GPU	nvidia	GeForce RTX 5090	1
Liquid Cooling	CPU	Hydro	1
Desktop case	Silent	Avlano	1
Disk 4 To	WD/Samsung	SN850/990 pro	2
RAM 192 Go	Corsair	48GB	4
Power supply	Alvano	1200W	1
Others	Screen/mouse/ keyboard		

**ATTENTION** hardware configuration indicative only, will be confirmed when order is received.  
**2U rack server** configurations for data centers also available (transtec.ch), 96GB (tox2) of GPU RTX 6000. Configurations/proposal on request.



# Software overview

Realtime technologies, XMPP, Jitsi, webRTC

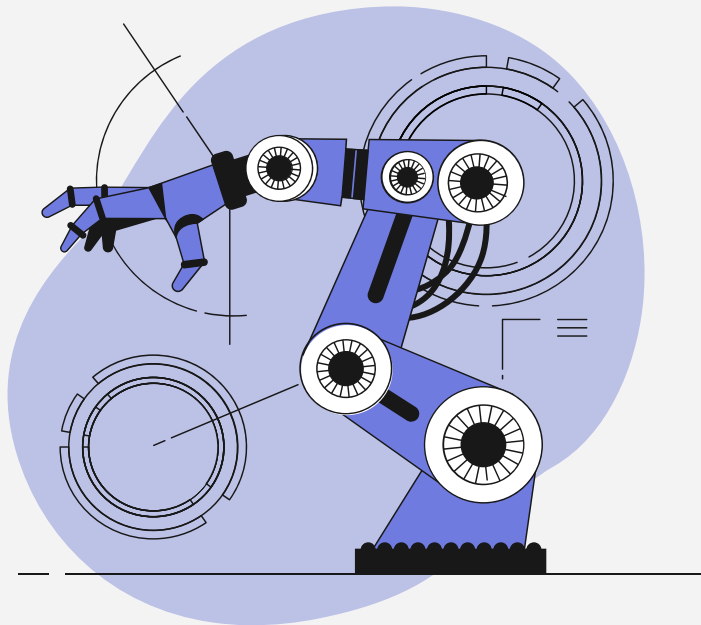
**igniterealtime.org**

LLM expertise: Huggingface, tuning and adjustments, prompting (Llama3, Mistral, OpenAI, Qwen, Lucie, Apertus, DeepSeek...).

C programs for RAG: Internet, DB, webRTC, AI, vocal bridges...read **pdf and images (Mistral)**.

And many other FREE expertise, ffmpeg, curl etc...

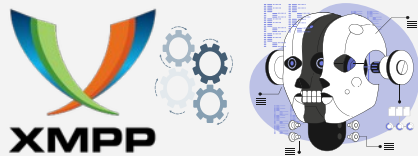




# Software details

- Kernel modules GPU : GNU
- OS : Linux Ubuntu/Debian
- Containers : Linux + Docker
- llama.cpp : LLM server
- Librechat (IA webchat) PC/mobile
- Text 2 Speech : gpu docker servers
- Speech 2 Text : Linto gpu servers
- Openfire and jitsi
- Progressive mobile application
- Integration with Chat XMPP & conversation, Jitsi...

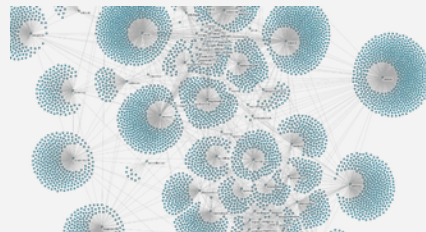




# XMPP + AI benefits



- AI is a chatbot, just add it to your contacts;
- Your chatbot 🦙 can be a member of a Jitsi visioconference;
- Your chatbot can interact by voice 🗣️ and be a member of an audio conference;
- Your chatbot knows how to answer a phone call 📞 JITSI+SIP+AI;
- Your chatbot can have multiple connexions together with XMPP Fediverse neural Internet network.

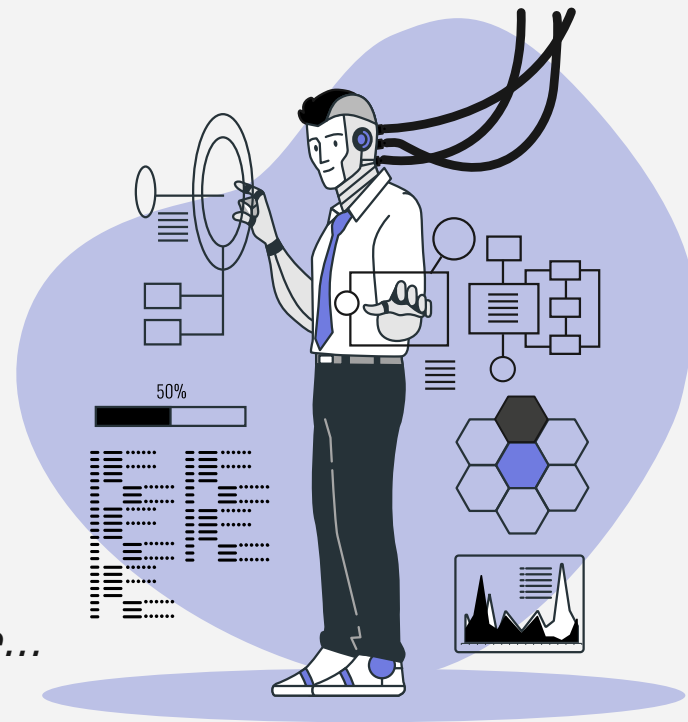


*Distribute and control your AI chatbots, both within the company  
and on the Internet.*

# Fonctions inventory

- ChatBot 🤖💬
- ChatBot VoIP 🤖💬☎️
- ChatBot Visioconference 🤖💬📹
- ChatBot Audioconference 🤖💬👤
- Multilingual translations 🇬🇧🇩🇪🇮🇹🇫🇷🇪🇸 (!)
- Questions ? Answers 🧠💡
- Self-discussion 🐐🗣️👂🐐🧠🧑🐐👏

*Read pdf and images, prompt, API and more...  
all you need for integration.*





# Reasons why it's so damn critical

- All your data stays on premise;
- Transcriptions generated on your infrastructure, neither on Google nor on Apple;
- Experiment all your expectations, with complete peace of mind, even with confidential information;
- Instantaneous access to ultra-fast voice and AI “**0-latency**”, thanks to local and integrated visioconference & webRTC;
- Immediately AI available as a simple chat contact and as participant in a visioconference, thank's XMPP.



*Everything that works, developed and validated on your aiborg desktop can be deployed very quickly in datacenter + clusters + HA*










# In action...

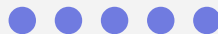
**aiborg** by appointment on <https://aiborg.free-solutions.ch> (voice)

<https://aiborg.free-solutions.ch:8443/aiborg> (visio)



<https://chatgpt.free-solutions.ch/> (prompt)

- [https://green-spider.free-solutions.org/AI\\_VIDEOS/](https://green-spider.free-solutions.org/AI_VIDEOS/) Videos
- Visioconferences & private chat  <https://www.free-solutions.org>
- Collaborative drives, zero trust, cryptpad  <https://cryptpad.free-solutions.ch>
- LINUX  replacement OS for laptops and desktops  
[https://github.com/ClaudeStabile/GREEN\\_SPIDER](https://github.com/ClaudeStabile/GREEN_SPIDER) 
- Mastodon servers, private enterprise instances  
<https://mastodon.free-solutions.org>  & <https://toot.igniterealtime.org>








# In case of a cyber attack

(shutdown of your IT)



- A complete, ready-to-use crisis communication solution:
  - Visio and chat platform;
  - Ready-to-use accounts for key players (open to the outside world via a dedicated Internet connection on DMZ);
  - Help of your trained AI for this mission.
- With Cryptpad\*, edited by 
- Ability to store and organize all your “cyber crisis” documents;
- Ability to collaborate and write in a dedicated, extremely secure office environment.



\* Cryptpad is an opensource software published by [XWIKI SAS](#) free-solutions has been integrating it for over 10 years. It can be installed as an option, activated or not, commissioning and training on request (project mode).



CryptPad




# Investment

20k€/kCHF VAT excluded



# AIBORG

- Delivered/billed from 
  - Hardware (approx. 8k€), delivery time, installation, commissioning, from 4 to 5 weeks.
  - Services (12k€), installations, tests, OS environment, virtualization and software, validations (includes 3 days of skills transfer for administrators, developers and project managers).

**ATTENTION** hardware configuration will be confirmed when order is received.  
**1U rack server** configurations for data centers also available ([transtec.ch](https://transtec.ch)), 48 or 96GB of GPU.  
Configurations on request.



# Thank you

## Questions?

contact@aiborg.eu



# AIBORG



your on premise AI packaged solution